

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-296502

(43)Date of publication of application : 29.10.1999

(51)Int.CI. G06F 17/00
G06F 17/10

(21)Application number : 10-102673 (71)Applicant : SHIRAYURI GAKUEN
HORII KIYOKI

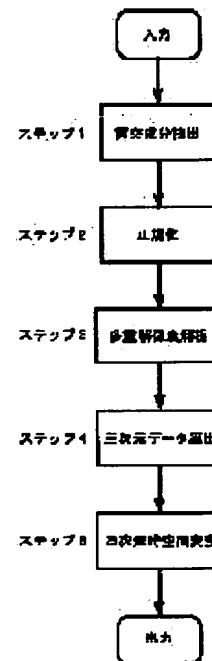
(22)Date of filing : 14.04.1998 (72)Inventor : SAITO CHIYOUKO
HORII KIYOKI
MIYAZAWA KENJI

(54) METHOD AND DEVICE FOR ANALYZING LITERATURE

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a completely new analyzing method and analyzing device for literature which can objectively analyze the sentence style structure of very subjective literature by using computing equipment such as a computer and a spatial theory method which is a mathematical method as a well-worn device and was not seen before.

SOLUTION: This is a method for analyzing literature, and three kinds of linear independent element data that the literature has, are subjected to an orthogonal component extracting process (S1) using Gram-Schmidt relation, a normalizing process (S2), and a multi-resolution analyzing process (S3) using wavelet conversion; and three-dimensional data including the obtained three kinds of normalized orthogonal system linear independent element data as (x)-, (y)-, and (z)-directional components are obtained at each resolution level (S4) and represented in a four-dimensional space;



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

BEST AVAILABLE COPY

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japan Patent Office

(19) 日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-296502

(43) 公開日 平成11年(1999)10月29日

(51) Int.Cl.⁶
G 0 6 F 17/00
17/10

識別記号

F I
G 0 6 F 15/20
15/31

Z
Z

審査請求 未請求 請求項の数4 O.L. (全22頁)

(21) 出願番号

特願平10-102673

(22) 出願日

平成10年(1998)4月14日

特許法第30条第1項適用申請有り 1998年3月31日 法政大学計算科学研究センター発行の「法政大学計算科学研究センター研究報告 第11巻」に発表

(71) 出願人 598049584

学校法人 白百合学園

東京都千代田区九段北二丁目4番1号

(71) 出願人 390031853

堀井 清之

東京都目黒区上目黒5丁目8番15-501号

(72) 発明者 斎藤 兆古

東京都あきる野市山田778-12

(72) 発明者 堀井 清之

東京都目黒区上目黒5-8-15-501

(72) 発明者 宮澤 賢治

山梨県甲府市飯田4-7-16

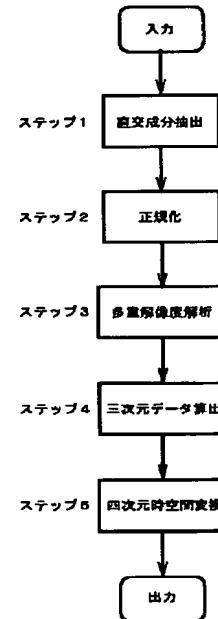
(74) 代理人 弁理士 西澤 利夫

(54) 【発明の名称】 文学作品解析方法および解析装置

(57) 【要約】

【課題】 主観的要素の強い文学作品の文体構造を、コンピュータ等の計算機器を用い、数学的手法である線形空間論的手法を常套手段として、客観的に解析することができる、従来にはない全く新しい文学作品の解析方法および解析装置を提供する。

【解決手段】 文学作品を解析する方法であって、文学作品が有する三種類の一次独立要素データそれぞれに対して、順次、グラムシュミット関係を用いた直交成分抽出処理、正規化処理、およびウェーブレット変換を用いた多重解像度解析処理を施し、各解像度レベル毎に、得られた三種類の正規直交系一次独立要素データそれぞれをx, y, z方向成分とした三次元データを求め、この三次元データを四次元時空間において表現する。



【特許請求の範囲】

【請求項1】 文学作品を解析する方法であって、文学作品が有する三種類の一次独立要素データそれぞれに対して、順次、グラムシュミット関係を用いた直交成分抽出処理、正規化処理、およびウェーブレット変換を用いた多重解像度解析処理を施し、各解像度レベル毎に、得られた三種類の正規直交系一次独立要素データそれぞれをx, y, z方向成分とした三次元データを求め、この三次元データを四次元時空間において表現することを特徴とする文学作品解析方法。

【請求項2】 三種類の一次独立要素データ各々からグラムシュミット関係を用いて直交成分を抽出し、各直交成分を正規化して正規直交系一次独立要素データを算出し、各正規直交系一次独立要素データ各々をウェーブレット変換を用いた多重解像度解析して各解像度レベル毎の各正規直交系一次独立要素データを算出し、各解像度レベル毎に、各正規直交系一次独立要素データそれぞれをx, y, z方向成分とした三次元データを算出し、各解像度レベル毎に、三次元データを四次元時空間に変換する請求項1の文学作品解析方法。

【請求項3】 文学作品を解析する装置であって、文学作品が有する三種類の一次独立要素データ各々からグラムシュミット関係を用いて直交成分を抽出する直交成分抽出処理手段と、各直交成分を正規化して正規直交系一次独立要素データを算出する正規化処理手段と、各正規直交系一次独立要素データをウェーブレット変換を用いた多重解像度解析して各解像度レベル毎の各正規直交系一次独立要素データを算出する多重解像度解析処理手段と、各解像度レベル毎に、各正規直交系一次独立要素データそれぞれをx, y, z方向成分とした三次元データを算出する三次元データ算出手段と、各解像度レベル毎に、三次元データを四次元時空間に変換する四次元時空間変換手段とを備えている文学作品解析装置。

【請求項4】 文学作品を解析する装置であって、文学作品が有する三種類の一次独立要素データを記憶する入力記憶手段と、記憶手段から各一次独立要素を読み出し各々からグラムシュミット関係を用いて直交成分を抽出する直交成分抽出処理手段と、各直交成分を記憶する直交成分記憶手段と、直交成分記憶手段から各直交成分を読み出し各々を正規化して正規直交系一次独立要素データを算出する正規化処理手段と、各正規直交系一次独立要素データを記憶する正規直交データ記憶手段と、

正規直交データ記憶手段から各正規直交系一次独立要素データを読み出しウェーブレット変換を用いた多重解像度解析して各解像度レベル毎の各正規直交系一次独立要素データを算出する多重解像度解析処理手段と、各解像度レベル毎の各正規直交系一次独立要素データを記憶する解像度レベルデータ記憶手段と、解像度レベルデータ記憶手段から各解像度レベル毎の各正規直交系一次独立要素データを読み出し、各解像度レベル毎に各正規直交系一次独立要素データそれぞれをx, y, z方向成分とした三次元データを算出する三次元データ算出手段と、各解像度レベル毎の三次元データを記憶する三次元データ記憶手段と、三次元データ記憶手段から各解像度レベル毎の三次元データを読み出し、各解像度レベル毎に三次元データを四次元時空間に変換する四次元時空間変換手段とを備えている文学作品解析装置。

【発明の詳細な説明】

【0001】

20 【発明の属する技術分野】この出願の発明は、文学作品解析方法および解析装置に関するものである。さらに詳しくは、この出願の発明は、コンピュータ等の計算機器を用いて、常套手段とした線形空間論的手法により文学作品を客観的に解析することのできる、従来ない全く新しい文学作品の解析方法および解析装置に関するものである。

【0002】

【従来の技術とその課題】一般に、文学作品は、視覚的表象である文字データを最小構成データとして構成されている。特定の意味を持つ文字データを組み合わせた単語データが二番目の構成データであり、この単語データが組み合わされて文章データが三番目の構成データとして構成され、そして、文章データが組み合わされて物語性を持つ文学作品が形成される。

30 【0003】従来より、このような構成によりなる文学作品の評価は、それが物語性、文章データの構成法、単語データの使われ方など、小さな構成データから大きな構成データ、さらには文学作品全体にわたる観点から主観的に行われている。この主観的評価は計数的になされることはなく、単語データ周辺の文章データや雰囲気等のような、計数的取り扱いが困難なバラメータを前提としてなされる。このため、評価を下す文学者等の知識、経験、考え方などによって異なる評価が行われている。

40 【0004】しかしながら、このような主観的な評価ではなく、文学作品の客観的な解析に関しては、コンピュータを用いて単語データを情報科学的に捉える研究などの他は、ほとんど研究・開発されておらず、数学的手法を常套手段とする計算科学的解析方法による客観性の高い解析は未だ実現されていない。そこで、この出願の発明は、以上の通りの事情に鑑みてなされたものであり、

主観的要素の強い文学作品の文体構造を、コンピュータ等の計算機器を用い、数学的手法である線形空間論的手法を常套手段として、客観的に解析することのできる、従来にはない全く新しい文学作品の解析方法および解析装置を提供することを目的としている。

【0005】

【課題を解決するための手段】この出願の発明は、上記の課題を解決するものとして、文学作品を解析する方法であって、文学作品が有する三種類の一次独立要素データそれぞれに対して、順次、グラムシュミット関係を用いた直交成分抽出処理、正規化処理、およびウェーブレット変換を用いた多重解像度解析処理を施し、各解像度レベル毎に、得られた三種類の正規直交系一次独立要素データそれぞれを x 、 y 、 z 方向成分とした三次元データを求め、この三次元データを四次元時空間において表現することを特徴とする文学作品解析方法（請求項1）を提供する。

【0006】また、この発明は、上記の方法において、三種類の一次独立要素データ各々からグラムシュミット関係を用いて直交成分を抽出し、各直交成分を正規化して正規直交系一次独立要素データを算出し、各正規直交系一次独立要素データ各々をウェーブレット変換を用いた多重解像度解析して各解像度レベル毎の各正規直交系一次独立要素データを算出し、各解像度レベル毎に、各正規直交系一次独立要素データそれぞれを x 、 y 、 z 方向成分とした三次元データを算出し、各解像度レベル毎に、三次元データを四次元時空間に変換すること（請求項2）をその態様として提供する。

【0007】さらにまた、この発明は、文学作品を解析する装置であって、文学作品が有する三種類の一次独立要素データ各々からグラムシュミット関係を用いて直交成分を抽出する直交成分抽出処理手段と、各直交成分を正規化して正規直交系一次独立要素データを算出する正規化処理手段と、各正規直交系一次独立要素データをウェーブレット変換を用いた多重解像度解析して各解像度レベル毎の各正規直交系一次独立要素データを算出する多重解像度解析処理手段と、各解像度レベル毎に、各正規直交系一次独立要素データそれぞれを x 、 y 、 z 方向成分とした三次元データを算出する三次元データ算出手段と、各解像度レベル毎に、三次元データを四次元時空間に変換する四次元時空間変換手段とを備えている文学作品解析装置（請求項3）、および文学作品を解析する装置であって、文学作品が有する三種類の一次独立要素データを記憶する入力記憶手段と、記憶手段から各一次独立要素を読み出し各々からグラムシュミット関係を用いて直交成分を抽出する直交成分抽出処理手段と、各直交成分を記憶する直交成分記憶手段と、直交成分記憶手段から各直交成分を読み出し各々を正規化して正規直交系一次独立要素データを算出する正規化処理手段と、各正規直交系一次独立要素データを記憶する正規直交データ記憶手段と、正規直交データ記憶手段から各正規直交系一次独立要素データを読み出しウェーブレット変換を用いた多重解像度解析して各解像度レベル毎の各正規直交系一次独立要素データを算出する多重解像度解析処理手段と、各解像度レベル毎の各正規直交系一次独立要素データを記憶する解像度レベルデータ記憶手段と、解像度レベルデータ記憶手段から各解像度レベル毎の各正規直交系一次独立要素データを読み出し、各解像度レベル毎に各正規直交系一次独立要素データそれぞれを x 、 y 、 z 方向成分とした三次元データを算出する三次元データ算出手段と、各解像度レベル毎の三次元データを記憶する三次元データ記憶手段と、三次元データ記憶手段から各解像度レベル毎の三次元データを読み出し、各解像度レベル毎に三次元データを四次元時空間に変換する四次元時空間変換手段とを備えている文学作品解析装置（請求項4）をも提供する。

【0008】

【発明の実施の形態】この発明の文学作品解析方法は、上記の通り、文学作品が有する三種類の一次独立要素データそれぞれに対して、順次、グラムシュミット関係を用いた直交成分抽出処理、正規化処理、およびウェーブレット変換を用いた多重解像度解析処理を施し、各解像度レベル毎に、得られた三種類の正規直交系一次独立要素データそれぞれを x 、 y 、 z 方向成分とした三次元データを求め、この三次元データを四次元時空間において表現するものである。

【0009】より詳しくは、図1にも例示したように、まず、入力された三種類の一次独立要素データをグラムシュミット関係に当てはめて、このグラムシュミット関係を計算することにより、各一次独立要素データからの直交成分を抽出し（ステップ1）、各直交成分に対して正規化処理を施すことにより正規直交系一次独立要素データを算出し（ステップ2）、算出された各正規直交系一次独立要素データ各々にウェーブレット変換を用いた多重解像度解析を施して各解像度レベル毎の各正規直交系一次独立要素データを算出し（ステップ3）、さらに、各解像度レベル毎の各正規直交系一次独立要素データを用いて、各解像度レベル毎に、各正規直交系一次独立要素データそれぞれを x 、 y 、 z 方向成分とした三次元データを算出し（ステップ4）、そして、各解像度レベル毎の三次元データを四次元時空間に変換して（ステップ5）、解像度レベル毎の三次元データを四次元時空間で出力表現する。このようにして、この発明により、コンピュータを用いて、主観的要素の強い文学作品の文体構造を、非常に容易に、且つ客観性高く解析することができる。以下、添付した図面に沿って実施例を示し、この発明の実施の形態についてさらに詳しく説明する。

【0010】

50 【実施例】（実施例1）この出願の発明の文学作品解析

方法により、コンピュータを用いて、夏目漱石による「虞美人草」における現代的思考を持つ人物（以下、現代型人物と呼ぶ）の頻出度合い、中間的思考、つまり現代的および古典的いずれにも属さない思考を持つ人物

（以下、中間型人物と呼ぶ）の頻出度合い、および古典的思考を持つ人物（以下、古典型人物と呼ぶ）の頻出度合いそれを、互いにかかわり合いのある三種類の一次独立要素データとして選択し、これら一次独立要素データである各人物の頻出度合いを客観的に解析する。

【0011】現代型人物の頻出度合いは、横軸に章番号をとり、縦軸に度合い（%）をとつてなる図2に例示した通りであり、頻出度合いの振幅は最初が大きく、中間の章でやや小さく、終わりに近づくと多少大きくなるが収束的となっている。また、中間型人物の頻出度合いは、図3に例示した通りであり、中間の10章および11章で振動的傾向にあることがわかる。

【0012】一方、古典型人物の頻出度合いは、図4に例示した通りとなり、中間の8章、9章および10章で変化が減少している特徴となっている。ここで、これら図2～図4を重ねて一括して現すと、図5に例示したようになる。この図5からは、各人物の頻出度合いにある程度の相関があることを知ることができる。

【0013】たとえば、現代型人物の頻出度合いベクトルデータ（以下、現代型人物ベクトルデータと呼ぶ）と中間型人物の頻出度合いベクトルデータ（以下、中間型人物ベクトルデータと呼ぶ）との間の角度は47.3837度であり、古典型人物の頻出度合いベクトルデータ（以下、古典型人物ベクトルデータと呼ぶ）と中間型人物ベクトルデータとの間の角度は27.3445である。従って、「虞美人草」文学作品では、古典型人物が中間型人物に近いものとして描かれていることとなる。また、現代型人物と古典型人物間の角度は53.8988となり、明らかに相反する人物として扱われている。

【0014】しかしながら、図5だけでは、このようにある程度の相関を知ることはできるものの、より詳細に、各思考タイプの人物が「虞美人草」全体においてどのように登場しているかを客観的に解析することは非常に困難である。そこで、この発明の方法を用いることによって、「虞美人草」における現代型人物、中間型人物および古典型人物それぞれの頻出度合いを、より詳しく客観的に解析する。

【0015】まず、前処理として、各ベクトルデータを、ゼロ要素を加えることにより2のべき乗の要素個数をもつベクトルデータに変形しておく。本実施例1では、各人物の頻出度合いベクトルデータの要素個数は19個があるので、13個のゼロ要素を最後尾に加えて32個の要素個数としておく。このような前処理で32個の要素個数を持つベクトルに変形された各人物の頻出度合いベクトルデータに対して、グラム・シュミット関係を用いた直交成分抽出処理を施す。

【0016】より具体的には、本実施例1では、グラム・シュミット関係を用い、現代型人物ベクトルデータおよび古典型人物ベクトルデータそれから、中間型人物ベクトルデータに直交する成分を抽出し、第一直交成分ベクトルデータおよび第二直交成分ベクトルデータを得る。これら第一直交成分ベクトルデータおよび第二直交成分ベクトルデータが正しく抽出されたものであるかを確認するために、各ベクトルデータと中間型人物ベクトルデータとの間の角度を算出したところ、算出角度は90度となったので、抽出処理が正しく行われたことがわかる。

【0017】また、第一直交成分ベクトルデータと第二直交成分ベクトルデータとの間の角度は、90度と算出されたので、第一直交成分ベクトルデータと第二直交成分ベクトルデータとは互いに直交していることが分かる。次いで、上記の第一直交成分ベクトルデータ、第二直交成分ベクトルデータ、および中間型人物ベクトルデータをそれぞれ正規化する。得られた各ベクトルデータは、それぞれ、正規直交系現代型人物ベクトルデータ、正規直交系古典型人物ベクトルデータ、および正規直交系中間型人物ベクトルデータと呼ぶこととする。

【0018】この正規化処理によって、前記直交成分抽出処理において発生した第一直交成分ベクトルデータ、第二直交成分ベクトルデータ、および中間型人物ベクトルデータそれの偏りが取り除かれることとなる。ここで、これら正規直交系ベクトルデータそれぞれを三次元空間のx、y、z方向成分とした三次元ベクトルデータを算出してみる。図6(a)は、この三次元ベクトルデータを例示したものである。この図6(a)において、軸方向xは正規直交系中間型人物ベクトルデータ、幅方向yは正規直交系現代型人物ベクトルデータ、高さ方向zは正規直交系古典型人物ベクトルデータとしている。また、図6(b)～(d)は、各々、正規直交系中間型人物ベクトルデータの大きさ、正規直交系現代型人物ベクトルデータの大きさ、正規直交系古典型人物ベクトルデータの大きさを例示したものである。

【0019】これら図6(a)～(d)から明らかにように、三次元ベクトルデータの大きさは、初期の章で小さく、中間の章で最大となり、終章に近づくに連れて減少していることがわかる。また、中間章では、単にベクトルの大きさが増加するだけでなく、現代型人物ベクトルデータおよび古典型人物ベクトルデータが交互に入れ代わって現れている。すなわち、現代的思考を持つ人物と古典的思考を持つ人物とが交互に現れて、「虞美人草」における山場を構成していることが、客観的に知ることができる。

【0020】しかしながら、図6(a)～(d)に例示した三次元ベクトルデータでは、まだ直交成分抽出処理において発生したバラツキが存在しており、このバラツキがより詳細で、正確な客観的解析を妨げる一要因とな

っている。そこで、この発明の解析方法では、このバツキを除去して、より詳細、且つ高精度な客観的解析を行うために、ウェーブレット変換を用いた多重解像度解析処理をさらに施す。

【0021】具体的には、まず、正規化処理後に得られた正規直交系中間型人物ベクトルデータ、正規直交系現代型人物ベクトルデータおよび正規直交系古典型人物ベクトルデータそれぞれをウェーブレット変換して、ウェーブレットスペクトラムデータを算出する。得られたウェーブレットスペクトラムデータは、それぞれ、中間型人物ウェーブレットスペクトラム、現代型人物ウェーブレットスペクトラム古典型人物ウェーブレットスペクトラムと呼ぶこととする。

【0022】次に、中間型人物ウェーブレットスペクトラム、現代型人物ウェーブレットスペクトラムおよび古典型人物ウェーブレットスペクトラムを、各解像度レベル毎に、逆ウェーブレット変換して多重解像度解析を行う。このようなウェーブレット変換および逆ウェーブレット変換による多重解像度解析処理により、上述したバツキを除去することができる。

【0023】そして、各解像度レベル毎の正規直交系中間型人物ベクトルデータ、正規直交系現代型人物ベクトルデータおよび正規直交系古典型人物ベクトルデータを用いて、それぞれをx、y、z方向成分とする三次元ベクトルデータを解像度レベル毎に作成して、各三次元ベクトルデータを四次元時空間で表現する。より具体的には、たとえば、各解像度レベル毎の正規直交系中間型人物ベクトルデータ、正規直交系現代型人物ベクトルデータおよび正規直交系古典型人物ベクトルデータは、それぞれ、各解像度レベルに従って、L1中間型人物ベクトルデータ、L2中間型人物ベクトルデータ・・・Ln中間型人物ベクトルデータ、L1現代型人物ベクトルデータ、L2現代型人物ベクトルデータ・・・Ln現代型人物ベクトルデータ、L1古典型人物ベクトルデータ、L2古典型人物ベクトルデータ・・・Ln古典型人物ベクトルデータと呼ぶこととする。ここでnは解像度レベルの個数を示す。

【0024】たとえば、解像度レベルを6個とした場合には、L1中間型人物ベクトルデータ、L2中間型人物ベクトルデータ・・・L6中間型人物ベクトルデータ、L1現代型人物ベクトルデータ、L2現代型人物ベクトルデータ・・・L6現代型人物ベクトルデータ、L1古典型人物ベクトルデータ、L2古典型人物ベクトルデータ・・・L6古典型人物ベクトルデータとなる。

【0025】そして、各解像度レベル毎に、L1・・・Ln中間型人物ベクトルデータ、1L1・・・Ln現代型人物ベクトルデータそれぞれを三次元空間のx、y、z方向成分とするL1・・・Ln三次元ベクトルデータを求める、各L1・・・Ln三次元ベクトルデータを、文学作品にお

ける物語の進行、たとえば章番号を時間軸とした四次元時空間に変換して表現する。

【0026】図7(a)～図12(a)は、それぞれ、多重解像度解析処理における解像度レベルを六つとした場合において得られた、解像度レベル1～解像度レベル6のL1～L6三次元ベクトルデータの四次元時空間表現を例示したものである。この四次元時空間表現における時間軸は章番号とした。また、参考として、図7(b)(c)(d)～図12(b)(c)(d)それぞれに、L1～L6三次元ベクトルデータx、y、z方向成分を例示した。

【0027】図7(a)～(d)に例示した解像度レベル1のL1三次元ベクトルデータの四次元時空間は、「虞美人草」全体の平均的な各人物の頻出度合いを表しており、この四次元時空間表現から明らかなように、「虞美人草」では、作品全体として中間型人物の登場で構成されていることがわかる。図8(a)～(d)に例示した解像度レベル2のL2三次元ベクトルデータの四次元時空間は、「虞美人草」を二分割した場合の各人物の頻出度合いが表しており、ほとんどの章は、中間型人物の登場で構成されており、終わりの三章では、全てのタイプの人物の頻出度材が減少している傾向を示している。

【0028】また、図9(a)～(d)に例示した解像度レベル3のL3三次元ベクトルデータの四次元時空間は、「虞美人草」を三分割した場合の各人物の頻出度合いを表しており、中間型人物の頻出度合いがほぼ単調に増加しており、現代型人物および古典型人物それぞれの頻出度合いはほぼ同じ傾向となっており、中間章で増加し、終章で減少していることがわかる。

【0029】図10(a)～(d)に例示した解像度レベル4のL4三次元ベクトルデータの四次元時空間は、「虞美人草」を五分割した場合の各人物の頻出度合いを表しており、中間型人物の頻出度合いは微小振動しながら増加する傾向にあり、現代型人物および古典型人物は、交互に現れ、最初から最後に向かって振動的に減少していることがわかる。

【0030】さらに、図11(a)～(d)に例示した解像度レベル5のL5三次元ベクトルデータの四次元時空間は、「虞美人草」の二章毎の各人物の頻出度合いを表しており、古典型人物の頻出度合いが支配的となっており、最初の1、2章は古典型人物の頻出度合いは低く、次の3、4章では急激に増加している。その後の5～8章では平均的な頻出度合いであり、9～16章では大きな増減を繰り返し、終わりの3章では、増減が低下して平均的な頻出度合いとなっていることが解析できている。

【0031】そして、図12(a)～(d)に例示した解像度レベル6のL6三次元ベクトルデータの四次元時空間表現によって、「虞美人草」の各章毎の各人物の頻

出度合いが解析され、現代型人物および古典型人物の頻出度合いが支配的である。現代型人物および古典型人物は、互いに相反する頻出度合いとなっており、中間後の章に向かって振動的に増加し、終章に地下鉄にしたがって急激に平均的頻出度合いとなっていることが解析できた。

【0032】以上のように、ウェーブレット変換を用いた、つまりウェーブレット変換および逆ウェーブレット変換による多重解像度解析によって、直交成分抽出処理におけるバラツキを低減させて、各解像度レベル毎に、「虞美人草」文学作品における、現代的思考を持つ人物、中間的思考を持つ人物および古典的思考を持つ人物それぞれの頻出度合いを、より詳細に、且つより容易に客観的解析することができる。

【0033】(実施例2) 本実施例2では、「虞美人草」から、相互に関係した3種類の一次独立要素データとして、客観話法、会話法、および、客観話法および会話法いずれにも属さない中間話法それぞれの頻出度合いを選択し、コンピュータを用いて、客観的解析を行う。

【0034】図13～図15は、各々、客観話法、中間話法および会話法の頻出度合いを例示したものであり、横軸は章番号、縦軸は頻出度合い(%)となっている。客観話法の頻出度合いは、図13に例示したように、その振幅が最初の頃は大きく、中間の章でやや小さく、終章に近づくと再び振動的に大きくなる。中間話法の頻出度合いは、図14に例示したように、中間章において振動振幅が小さくなるが、中間章以降で大きな振幅で振動し、終章で収束する傾向となっている。

【0035】会話法の頻出度合いは、図15に例示したように、その振動が終章で収束しないような傾向となっていることがわかる。また、これら図13～図15を一括して表すと、図16に例示した通りとなる。ここで、各話法の頻出度合い間の相関を求めるとき、たとえば、中間話法ベクトルデータと客観話法ベクトルデータ間の角度は41.1147、中間話法ベクトルデータと会話法ベクトルデータ間の角度は54.7439と算出することができる。さらに、客観話法ベクトルデータと会話法ベクトルデータ間の角度は35.6356となっている。よって、これらの相関角度から、客観話法ベクトルデータと会話法ベクトルデータとが最も近い指向ベクトルであることがわかる。

【0036】この発明の解析方法では、さらに詳しく各話法の頻出度合いの客観的解析を行うために、実施例1と同様にして、グラムシュミットを用いた直交成分抽出処理、正規化処理、およびウェーブレット変換を用いた多重解像度解析処理を順次施し、得られる三次元データを四次元時空間にて表現する。まず、前処理として、19個の要素からなる元の各話法ベクトルデータを、ゼロ要素を加えることにより2のべき乗である32個の要素からなるベクトルデータに変形する。

【0037】そして、各話法のベクトルデータに対してグラムシュミット関係を用いた直交成分抽出処理を施す。本実施例2では、まず、グラムシュミット関係を用いて客観話法ベクトルデータから中間話法ベクトルデータに直交する成分を抽出し、第一直交成分ベクトルデータを得る。また、会話法ベクトルデータから中間話法ベクトルデータに直交する成分を抽出して、第二直交成分ベクトルデータを得る。

【0038】第一直交成分ベクトルデータおよび第二直交成分ベクトルデータそれぞれと中間話法ベクトルデータとの間の角度は、それぞれ、90度となっているので、抽出処理が精度良く行われていることが確認できる。さらにまた、第一直交成分ベクトルデータおよび第二直交成分ベクトルデータとの間の角度は、90度となり、互いに直交していることがわかる。

【0039】次に、上記のように直交であることが判定された第一直交成分ベクトルデータ、第二直交成分ベクトルデータおよび中間話法ベクトルデータに対して正規化処理をほどこし、各々が有する直交成分抽出処理時に発生した偏りを取り除く。得られる各データは、それぞれ、正規直交系客観話法ベクトルデータ、正規直交系会話法ベクトルデータ、正規直交系中間話法ベクトルデータと呼ぶこととする。

【0040】ここで、本実施例2でも、上述の実施例1と同様に、これら各正規直交系ベクトルデータを三次元空間の各軸方向x、幅方向y、高さ方向z成分として用いた三次元ベクトルデータを求めたところ、図17(a)に例示した通りとなった。また、図17(b)(c)(d)は、各々、軸方向x成分である正規直交系中間話法ベクトルデータの大きさ、幅方向y成分である正規直交系客観話法ベクトルデータの大きさ、高さ方向z成分である正規直交系会話法ベクトルデータの大きさを例示したものである。

【0041】これら図17(a)～(d)から明らかのように、最初の章から中間の章までは中間話法が支配しており、後半の章から終章までは会話法が支配している。客観話法も終章に近づく程大きくなっているが、この「虞美人草」における山場は会話法で述べられていることが客観的に解析することができるものの、この三次元ベクトルデータにはやはり直交成分抽出処理の際に発生した各ベクトルデータのバラツキが存在している。

【0042】よって、実施例1と同様にして、さらに、ウェーブレット変換を用いた多重解像度解析処理を施す。正規直交系客観話法ベクトルデータ、正規直交系会話法ベクトルデータおよび正規直交系中間話法ベクトルデータそれぞれをウェーブレット変換して、客観話法ウェーブレットスペクトラム、会話法ウェーブレットスペクトラムおよび中間話法ウェーブレットスペクトラムデータを算出し、各ウェーブレットスペクトラムを逆ウェーブレット変換して多重解像度解析する。

11

【0043】そして、各解像度レベル $L_1 \dots L_n$ 毎に、得られた $L_1 \dots L_n$ 正規直交系中間型人物ベクトルデータ、 $L_1 \dots L_n$ 正規直交系現代型人物ベクトルデータおよび $L_1 \dots L_n$ 正規直交系古典型人物ベクトルデータそれぞれを x 、 y および z 方向成分とした $L_1 \dots L_n$ 三次元ベクトルデータを作成し、さらに四次元時空間で表現する。

【0044】本実施例2では、解像度レベルを六個として、 $L_1 \dots L_6$ 三次元ベクトルデータを四次元時空間表現した。この四次元時空間における時間軸は章番号としてある。図18 (b) ~ (d) に例示した解像度レベル1の L_1 三次元ベクトルデータ、つまり「虞美人草」全体の平均的な各話法の頻出度合いは、図18 (a) に例示した通りに四次元時空間で表現でき、この四次元時空間表現から、この作品は作品全体として中間話法と客観話法で構成されていることがわかる。

【0045】図19 (b) ~ (d) に例示した解像度レベル2の L_2 三次元ベクトルデータ、つまり「虞美人草」を二分割した場合の各話法の頻出度合いは、図19 (a) に例示した通りの四次元時空間表現となり、ほとんどの章は中間話法で構成されているが、終わりの3章では中間話法の頻出度合いが減少していることがわかる。

【0046】図20 (b) ~ (d) に例示した解像度レベル3の L_3 三次元ベクトルデータ、つまり「虞美人草」を三分割した場合の各話法の頻出度合いは、図20 (a) に例示した通りの四次元時空間表現となり、中間話法に代って客観話法が支配的になりつつある。客観話法は、最初多く、中間で減少し、終章に近づくと再び増加する傾向にある。

【0047】図21 (b) ~ (d) に例示した解像度レベル4の L_4 三次元ベクトルデータ、つまり「虞美人草」を五分割した場合の各話法の頻出度合いは、図21 (a) に例示した通りの四次元時空間表現となり、会話法が支配的になり、統いて中間話法が支配的である。会話法は中間の章以降で急激な増減振動を行い、終章で一定値に収束するようになっている。

【0048】図22 (b) ~ (d) に例示した解像度レベル5の L_5 三次元ベクトルデータ、つまり「虞美人草」の二章毎の各話法の頻出度合いは、図22 (a) に例示したような四次元時空間表現になり、会話法の頻出度合いが支配的であり、最初の1~4章では会話法の頻出度合いはやや小さく、次の5~8章ではさらに減少するが、その後の9~12章では増減の振幅が大きくなり、13~16章でさらに大きな増減を繰り返すようになり、そして終わりの17~19章では会話法が減少し、中間話法の増減が支配的となる。これは13~16章が、この作品の山場を形成していることを意味し、上述した実施例1における思考タイプ別人物解析での古典型人物の傾向と類似している。

12

【0049】そして、図23 (b) ~ (d) に例示した最後の解像度レベル6の L_6 三次元ベクトルデータ、つまり「虞美人草」の各章毎の各話法の頻出度合いは、図23 (a) に例示したように四次元時空間において表現され、この四次元時空間表現から、客観話法の頻出度合いが支配的であり、中間章以降に大きな振動的変化があるが、終章に近づくと収束する傾向となっていることが客観的に解析できる。

【0050】以上の実施例1および実施例2においてこの発明の文学作品解析方法によって夏目漱石の「虞美人草」を客観的に解析した結果、登場人物の思考タイプ別分析では、現代型人物と古典型人物とが交互に登場し、その頻出度合いは中間の章に向かって振動的に増加し、その後終章に向かって平均的頻出度合いとなることが解析でき、また、話法別分析では、会話法が作品の山場を構成していることが解析できた。この会話法および古典型人物とは同じ傾向の頻出度合いを有していることも判明した。

【0051】なお、この発明の解析方法におけるウェーブレット変換および逆ウェーブレット変換に用いる基底関数としては、たとえばDaubechies (ドビッキー) の二次基底やHaar (ハール) 基底を用いることができる。上述した実施例1および実施例2では、ドビッキーの二次基底を用いている。(実施例3) 上述したこの発明の方法による文学作品の解析は、たとえば、図24に例示したような解析装置により行うことができる。

【0052】図24に例示したこの発明の文学作品解析装置は、直交成分抽出処理手段(1)、正規化処理手段(2)、多重解像度解析処理手段(3)、三次元データ算出手段(4)および四次元空間変換手段(5)を備えている。直交成分抽出処理手段(1)により、たとえば入力装置(図示していない)から入力された三種類の一次独立要素データ、たとえば上述した思考タイプ別の人々の頻出度合いベクトルデータなどがそれぞれ選択されてグラムシュミット関係に当てはめられ、このグラムシュミット関係が計算されて、各一次独立要素データからの直交成分が抽出される。

【0053】直交成分抽出処理手段(1)から受け取った直交成分には、それぞれ、正規化処理手段(2)により、正規化処理が施されて、各正規直交系一次独立要素データが得られる。次いで、多重解像度解析処理手段(3)により、正規化処理手段(2)から受け取った各正規直交系一次独立要素データに、上述した方法における処理と同様にして、ウェーブレット変換を用いた多重解像度解析処理が施されて、解像度レベル $L_1 \dots L_n$ に多重解析された三種類それぞれの $L_1 \dots L_n$ 正規直交系一次独立要素データが算出される。

【0054】さらに、三次元データ算出手段(4)により、多重解像度解析処理手段(3)から送られてきた各解像度レベル毎の各正規直交系一次独立要素データによ

50

って、各解像度レベル $L_1 \dots L_n$ 毎に、三種類の $L_1 \dots L_n$ 正規直交系一次独立要素データそれを x, y, z 方向成分とした $L_1 \dots L_n$ 三次元データが求められる。

【0055】そして、四次元時空間変換手段（5）により、三次元データ算出手段（4）からの $L_1 \dots L_n$ 三次元データが、文学作品における章番号などの物語進行を時間軸とした四次元時空間に変換されて表現される。このようにしてこの発明の解析装置により得られた四次元時空間表現が、たとえばCRT等の表示装置によって表示され、これを用いて、実施例1および実施例2で述べたように、文学作品を容易に客観的解析することができる。

【0056】図25は、この発明の文学作品解析装置の別の一例を示した構成図である。この図25に例示したこの発明の文学作品解析装置は、図24の文学作品解析装置における各手段に加え、各データを記憶する各種記憶手段を備えている。より詳しくは、入力装置（図示していない）から入力された三種類の一次独立要素データは、まず入力記憶手段（11）に記憶される。

【0057】直交成分抽出処理手段（1）は、この入力記憶手段（11）から三種類の一次独立要素データをそれぞれ選択的に読み出し、グラムシュミット関係に当てはめる。そして、このグラムシュミット関係を計算して各一次独立要素データから直交成分を抽出する。抽出された各直交成分は、直交成分記憶手段（12）に送られて記憶される。

【0058】正規化処理手段（2）は、直交成分記憶手段（12）から各直交成分を読み出し、各々に正規化処理を施して、各正規直交系一次独立要素データを算出し、これを記憶のために正規直交データ記憶手段（13）に送る。次いで、多重解像度解析処理手段（3）は、正規直交データ記憶手段（13）から読み出した各正規直交系一次独立要素データに、上述した方法における処理と同様にして、ウェーブレット変換を用いた多重解像度解析処理を施し、解像度レベル $L_1 \dots L_n$ に多重解析された三種類それを $L_1 \dots L_n$ 正規直交系一次独立要素データを算出する。

【0059】これら各解像度レベル毎の各正規直交系一次独立要素データはそれぞれ、解像度レベルデータ記憶手段（14）に送られて記憶される。さらに、三次元データ算出手段（4）は、解像度レベルデータ記憶手段（14）から読み出した各解像度レベル毎の各正規直交系一次独立要素データを用いて、各解像度レベル $L_1 \dots L_n$ 毎に、三種類の $L_1 \dots L_n$ 正規直交系一次独立要素データそれを x, y, z 方向成分とした $L_1 \dots L_n$ 三次元データを作成する。

【0060】作成された各三次元データは三次元データ記憶手段（15）に記憶される。そして、四次元時空間変換手段（5）が、三次元データ記憶手段（15）から

読み出した $L_1 \dots L_n$ 三次元データを、文学作品における章番号などの物語進行を時間軸とした四次元時空間に変換する。このようにして得られた四次元時空間における三次元データが、たとえば出力記憶手段（16）に記憶されるなどした後、出力表現されることとなる。

【0061】このように各手段により得られた各種データを、各記憶手段に記憶しておくことで、後に、各処理・算出・変換後のデータを表示や出力させる場合や、他の装置等によるデータとを比較したりする場合などに、非常に便利であり、コンピュータのハードウェア資源をより有効に活用している。これに対し、図24の解析装置では、記憶手段に記憶させる動作や記憶手段から読み出す動作などが省かれているので、より高速に解析処理を完了させることができる。

【0062】なお、図24および図25に例示したこの発明の装置は、たとえば、ウェーブレット変換および逆ウェーブレット変換それぞれに用いられるウェーブレット変換行列および逆ウェーブレット変換行列それぞれを一種類または複数種類あらかじめ記憶した変換行列記憶手段（図示していない）を備えている。この場合に、多重解像度解析処理手段（3）は、各々、変換行列記憶手段から選択的にウェーブレット変換行列および逆ウェーブレット変換行列を読み出して、ウェーブレット変換および逆ウェーブレット変換を行う。

【0063】また、あらかじめ一種類または複数種類の基底関数を記憶した基底関数記憶手段（図示していない）を備えている。この場合には、多重解像度解析処理手段は、基底関数記憶手段から選択的に基底関数を読み出し、この基底関数を用いてウェーブレット変換行列および逆ウェーブレット変換行列を作成して、ウェーブレット変換および逆ウェーブレット変換を行う。

【0064】これらのウェーブレット変換行列あるいは基底関数の選択では、キーボードやテンキーやマウス等の入力手段などからの入力信号により所望のものを選択することができる。もちろん、この発明は以上の例に限定されるものではなく、細部については様々な態様が可能であることは言うまでもない。

【0065】

【発明の効果】以上詳しく説明した通り、この発明の文学作品解析方法および解析装置によって、主観的要素の強い文学作品の文体構造を、非常に容易に、且つ客観性高く解析することができ、コンピュータ等の計算機器を用い、数学的手法を常套手段とする計算科学的方法に従った文学作品の客観的解析という全く新しい分野を開拓することができる。

【図面の簡単な説明】

【図1】この発明の文学作品解析方法を例示したフロー図である。

【図2】現代型人物の頻出度合いを例示した図である。

【図3】中間型人物の頻出度合いを例示した図である。

50

【図4】古典型人物の頻出度合いを例示した図である。

【図5】現代型人物、中間型人物および古典型人物それぞれの頻出度合いを重ね表示した一例を示した図である。

【図6】(a) (b) (c) (d)は、各々、ウェーブレット変換を用いた多重解像度解析処理前の三次元ベクトルデータ、およびこの三次元ベクトルデータのx, y, z方向成分を例示した図である。

【図7】(a) (b) (c) (d)は、各々、解像度レベル1のL1三次元ベクトルデータの四次元空間表現、およびL1三次元ベクトルデータのx, y, z方向成分を例示した図である。

【図8】(a) (b) (c) (d)は、各々、解像度レベル2のL2三次元ベクトルデータの四次元空間表現、およびL2三次元ベクトルデータのx, y, z方向成分を例示した図である。

【図9】(a) (b) (c) (d)は、各々、解像度レベル3のL3三次元ベクトルデータの四次元空間表現、およびL3三次元ベクトルデータのx, y, z方向成分を例示した図である。

【図10】(a) (b) (c) (d)は、各々、解像度レベル4のL4三次元ベクトルデータの四次元空間表現、およびL4三次元ベクトルデータのx, y, z方向成分を例示した図である。

【図11】(a) (b) (c) (d)は、各々、解像度レベル5のL5三次元ベクトルデータの四次元空間表現、およびL5三次元ベクトルデータのx, y, z方向成分を例示した図である。

【図12】(a) (b) (c) (d)は、各々、解像度レベル6のL6三次元ベクトルデータの四次元空間表現、およびL6三次元ベクトルデータのx, y, z方向成分を例示した図である。

【図13】客観話法の頻出度合いを例示した図である。

【図14】中間話法の頻出度合いを例示した図である。

【図15】会話法の頻出度合いを例示した図である。

【図16】客観話法、中間話法、会話法それぞれの頻出度合いを重ね表示した一例を示したものである。

【図17】(a) (b) (c) (d)は、各々、ウェーブレット変換を用いた多重解像度解析処理前の三次元ベクトルデータ、およびこの三次元ベクトルデータのx, y, z方向成分を例示した図である。

* 【図18】(a) (b) (c) (d)は、各々、解像度レベル1のL1三次元ベクトルデータの四次元空間表現、およびL1三次元ベクトルデータのx, y, z方向成分を例示した図である。

【図19】(a) (b) (c) (d)は、各々、解像度レベル2のL2三次元ベクトルデータの四次元空間表現、およびL2三次元ベクトルデータのx, y, z方向成分を例示した図である。

【図20】(a) (b) (c) (d)は、各々、解像度レベル3のL3三次元ベクトルデータの四次元空間表現、およびL3三次元ベクトルデータのx, y, z方向成分を例示した図である。

【図21】(a) (b) (c) (d)は、各々、解像度レベル4のL4三次元ベクトルデータの四次元空間表現、およびL4三次元ベクトルデータのx, y, z方向成分を例示した図である。

【図22】(a) (b) (c) (d)は、各々、解像度レベル5のL5三次元ベクトルデータの四次元空間表現、およびL5三次元ベクトルデータのx, y, z方向成分を例示した図である。

【図23】(a) (b) (c) (d)は、各々、解像度レベル6のL6三次元ベクトルデータの四次元空間表現、およびL6三次元ベクトルデータのx, y, z方向成分を例示した図である。

【図24】この発明の文学作品解析装置の一例を示した要部構成図である。

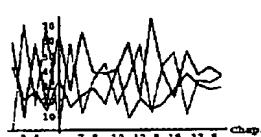
【図25】この発明の文学作品解析装置の別の一例を示した要部構成図である。

【符号の説明】

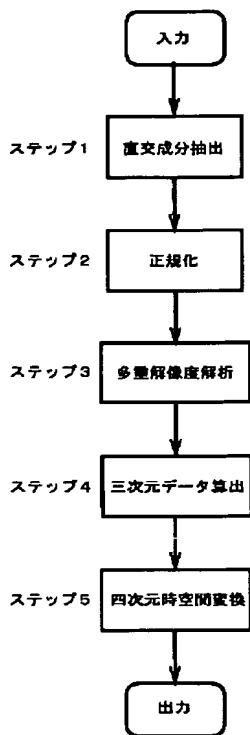
- 30 1 直交成分抽出処理手段
- 2 正規化処理手段
- 3 多重解像度解析処理手段
- 4 三次元データ算出手段
- 5 四次元時空間変換手段
- 11 入力記憶手段
- 12 直交成分記憶手段
- 13 正規直交データ記憶手段
- 14 解像度レベルデータ記憶手段
- 15 三次元データ記憶手段
- 40 16 出力記憶手段

*

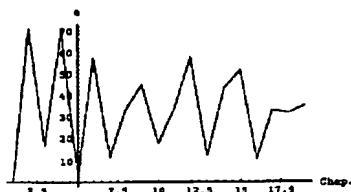
【図5】



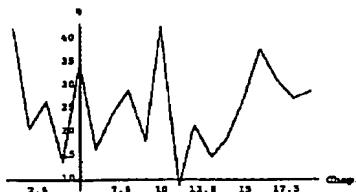
【図1】



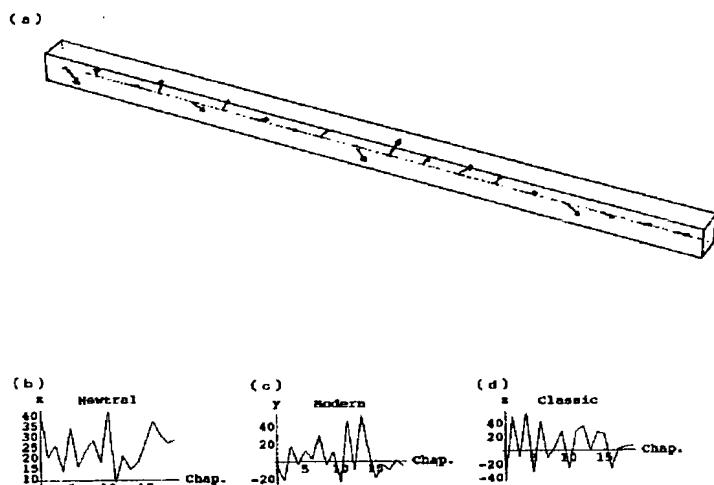
【図2】



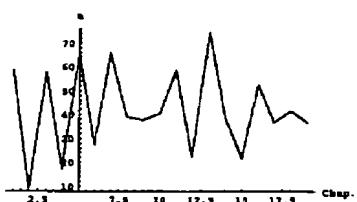
【図3】



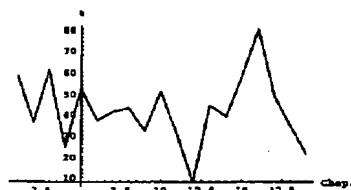
【図6】



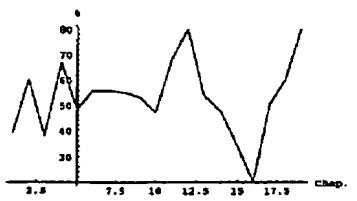
【図4】



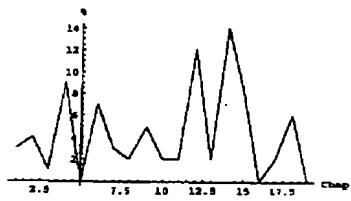
【図15】



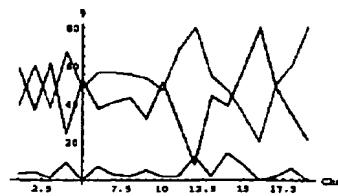
【図13】



【図14】

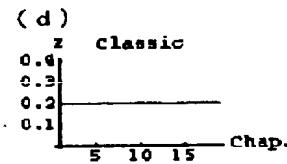
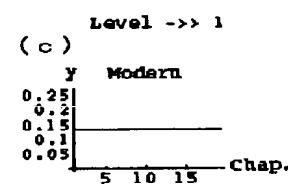
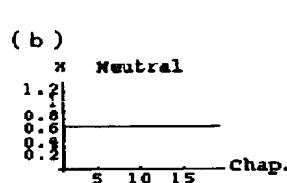
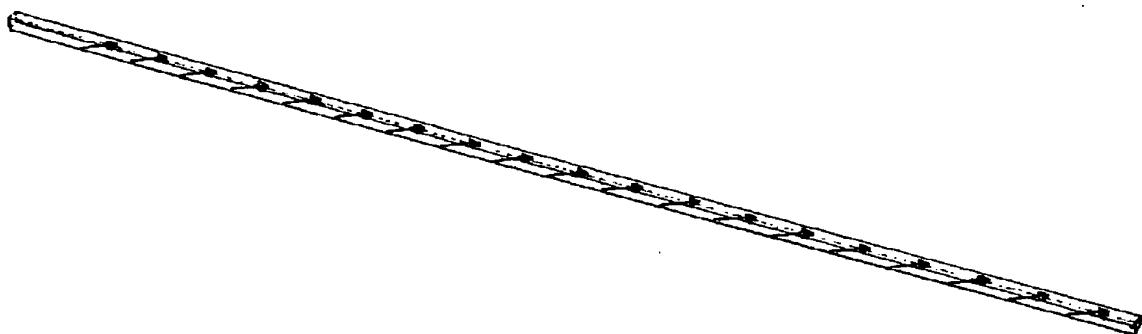


【図16】



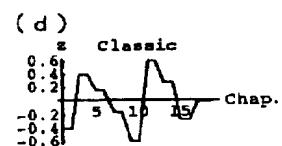
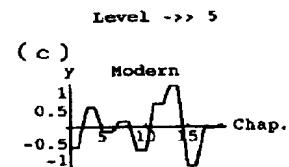
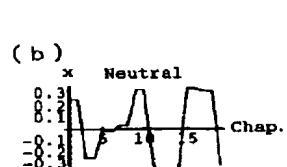
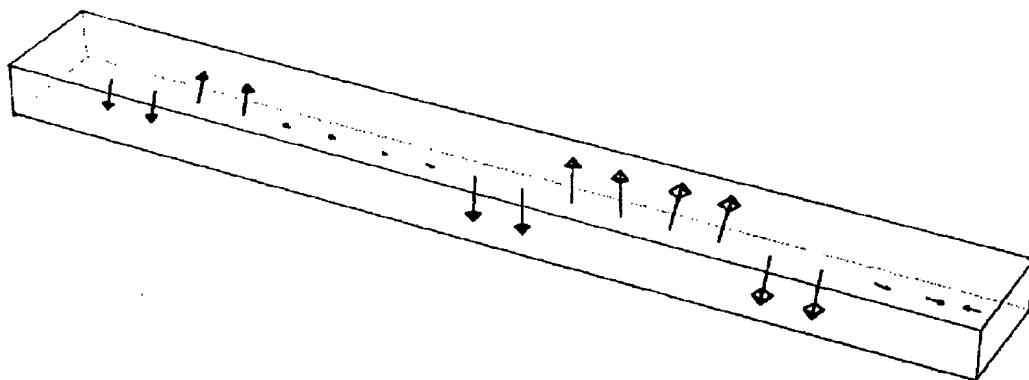
【図7】

(a)



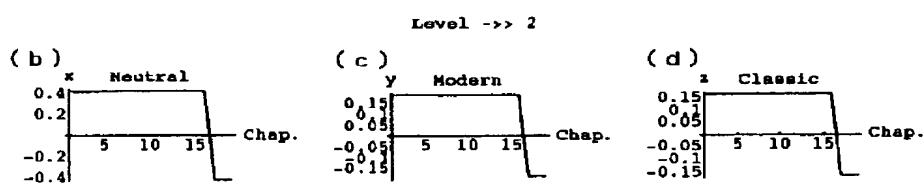
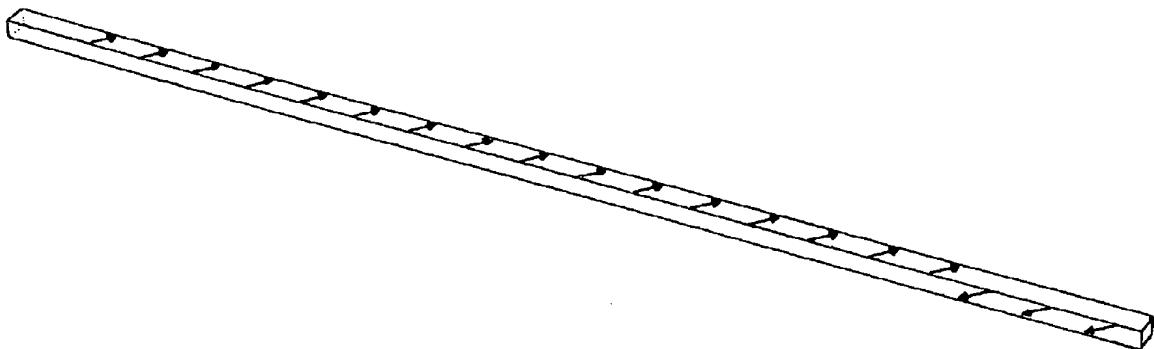
【図11】

(a)

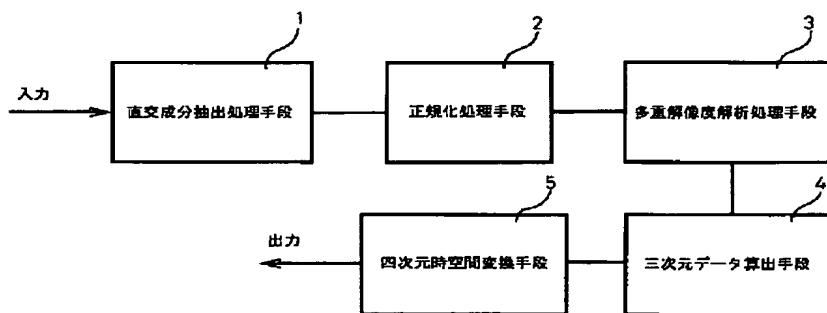


【図8】

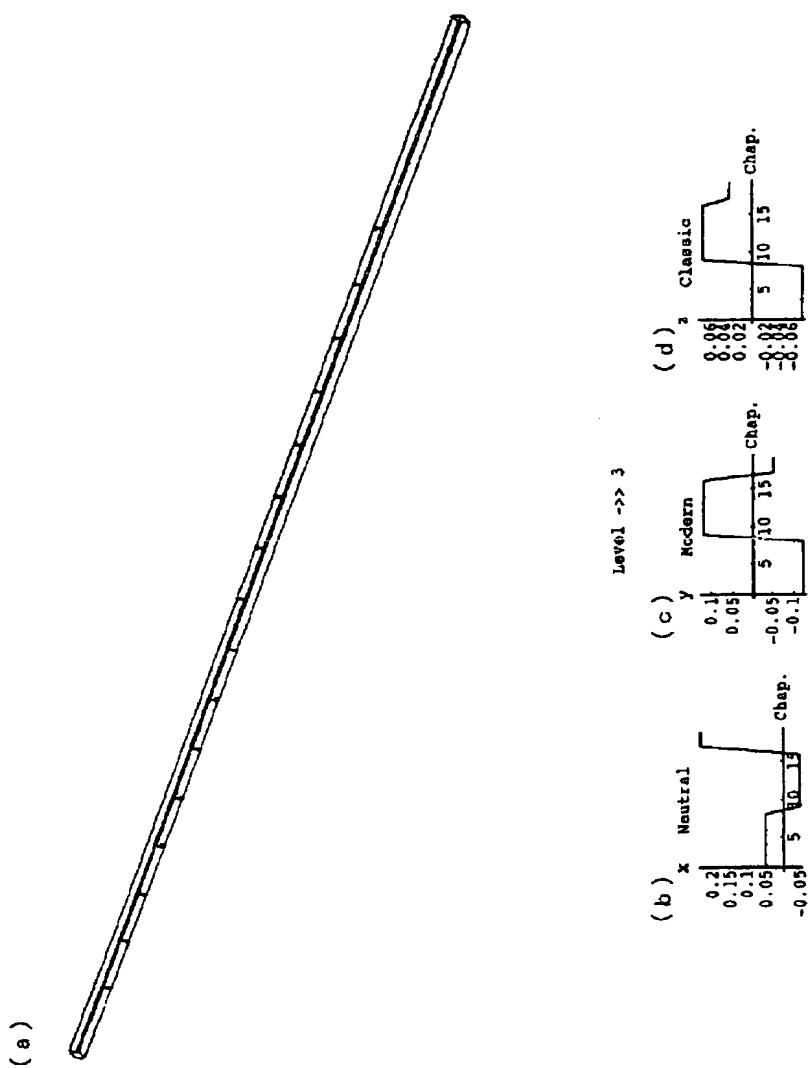
(a)



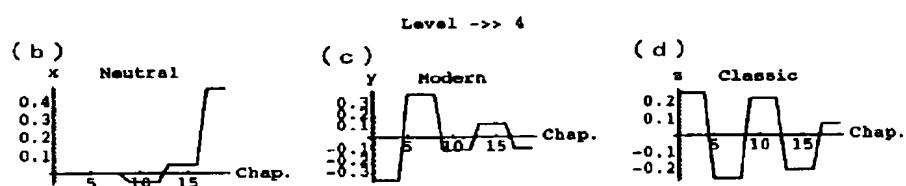
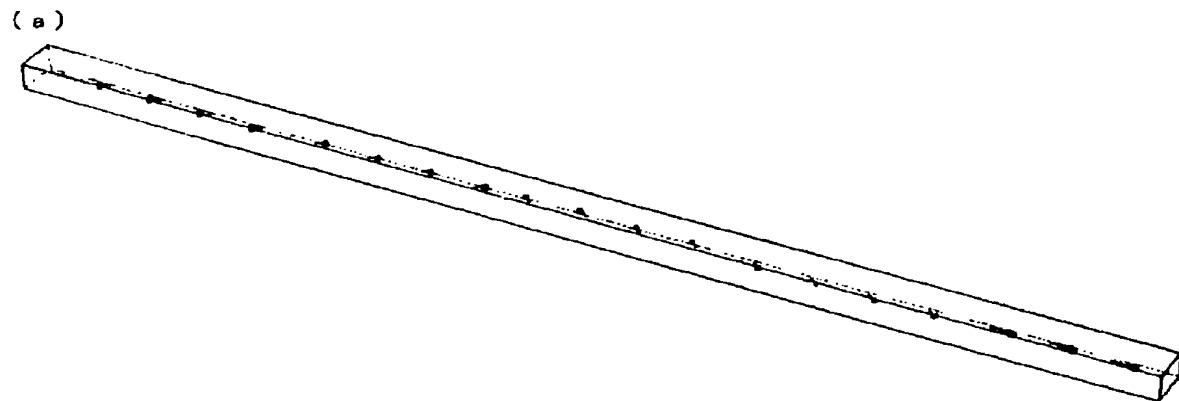
【図24】



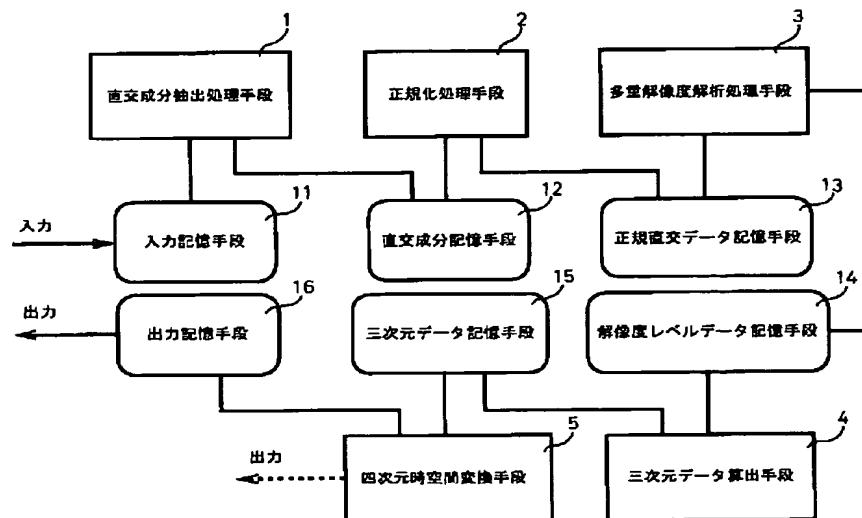
【図9】



【図10】

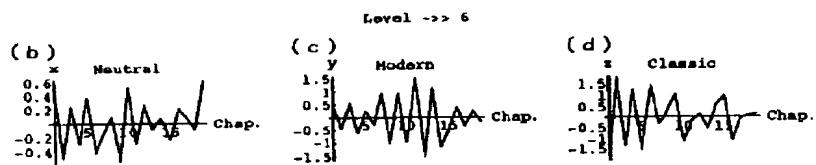
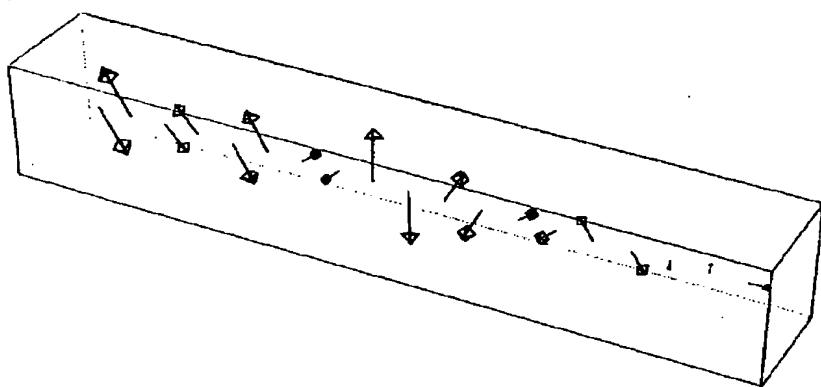


【図25】



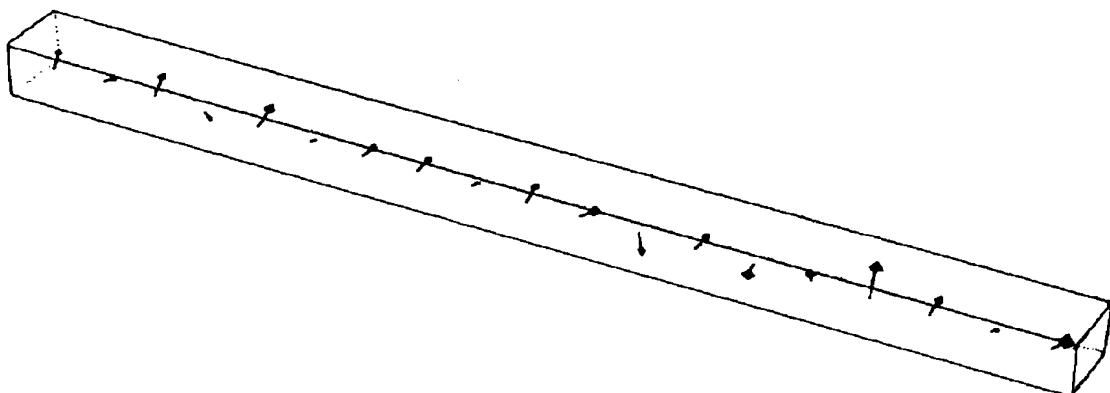
【図12】

(a)

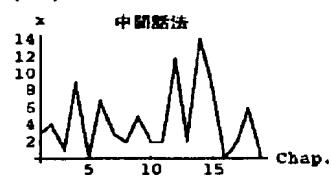


【図17】

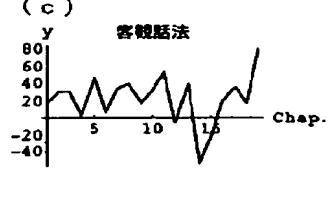
(a)



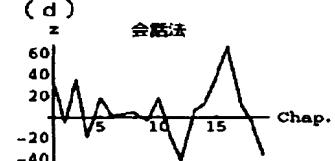
(b)



(c)

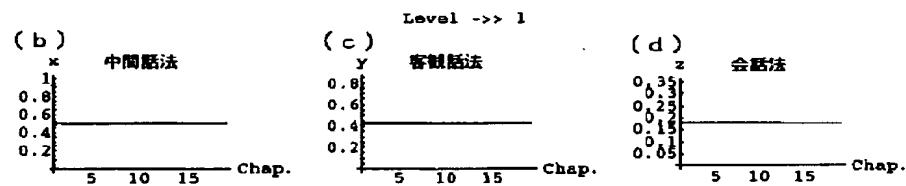
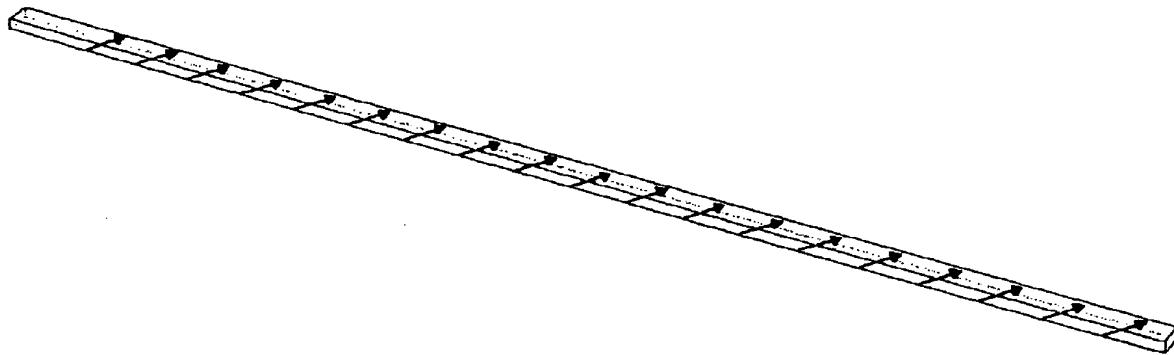


(d)



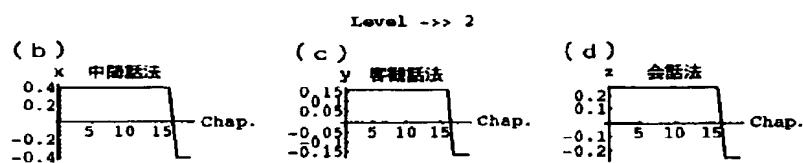
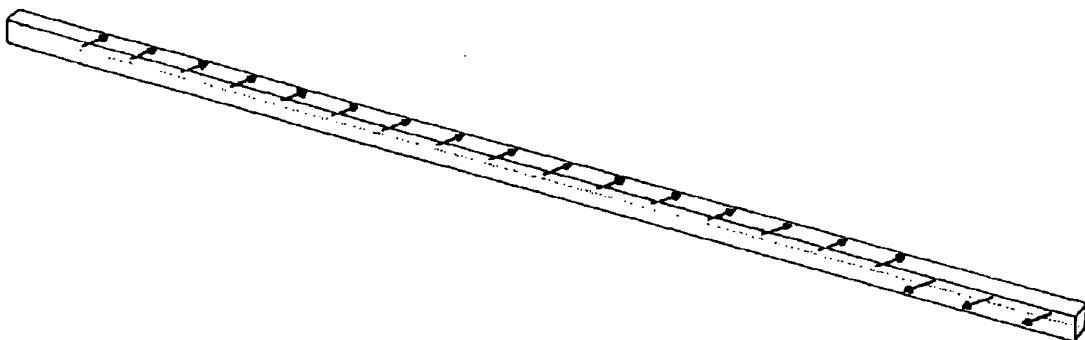
〔図18〕

(a)

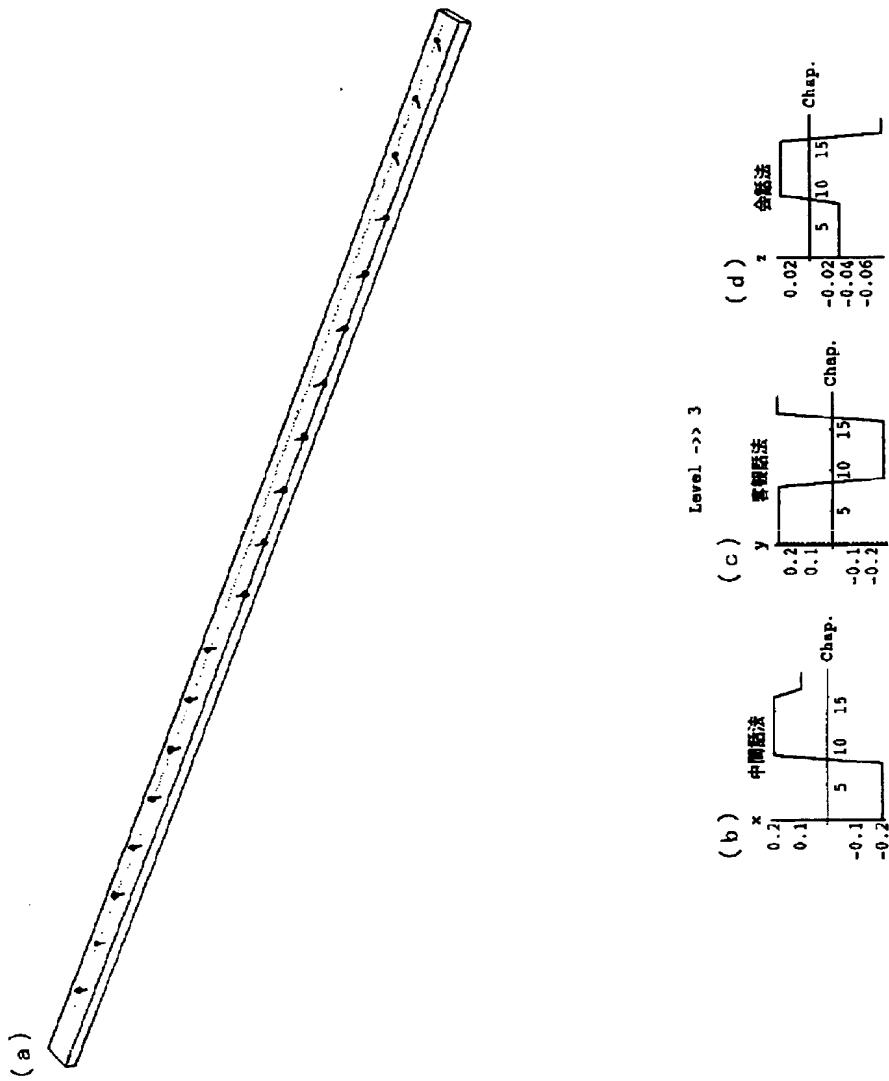


[図19]

(a)

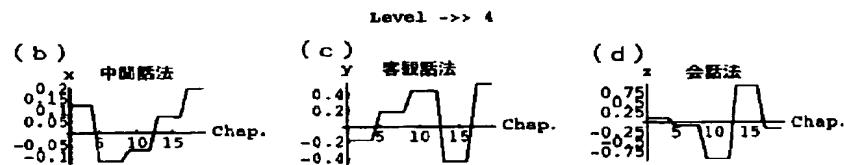
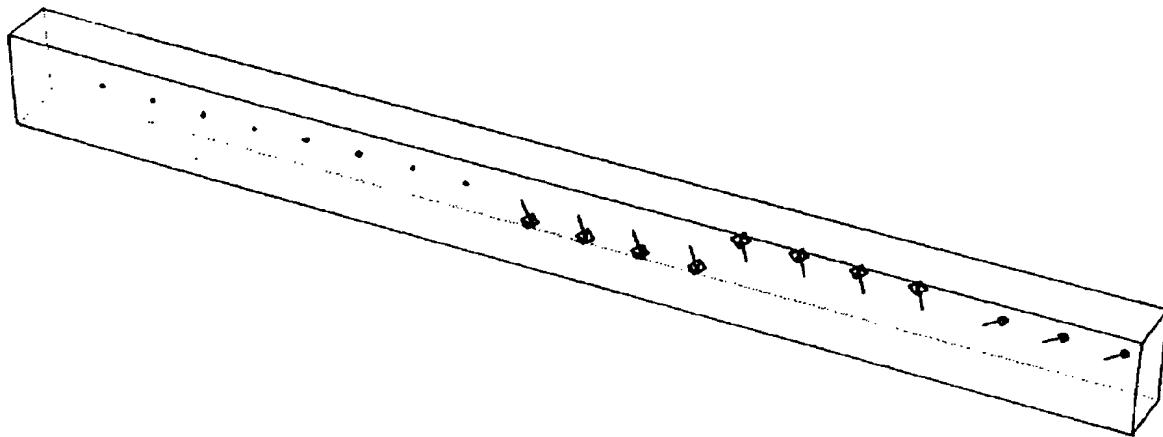


〔図20〕



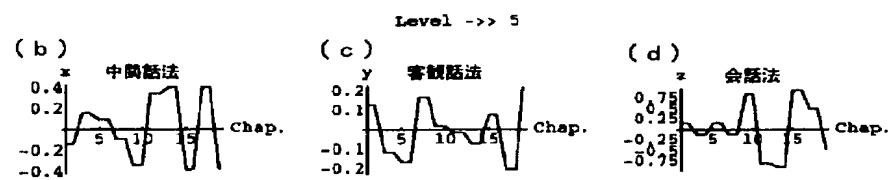
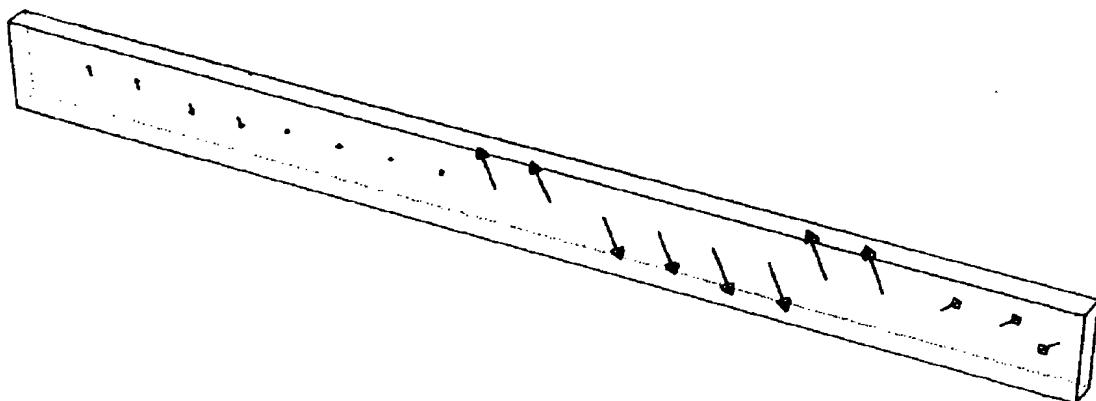
【図21】

(a)



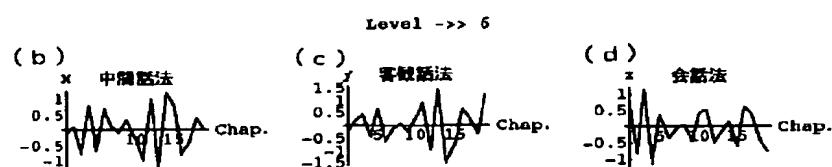
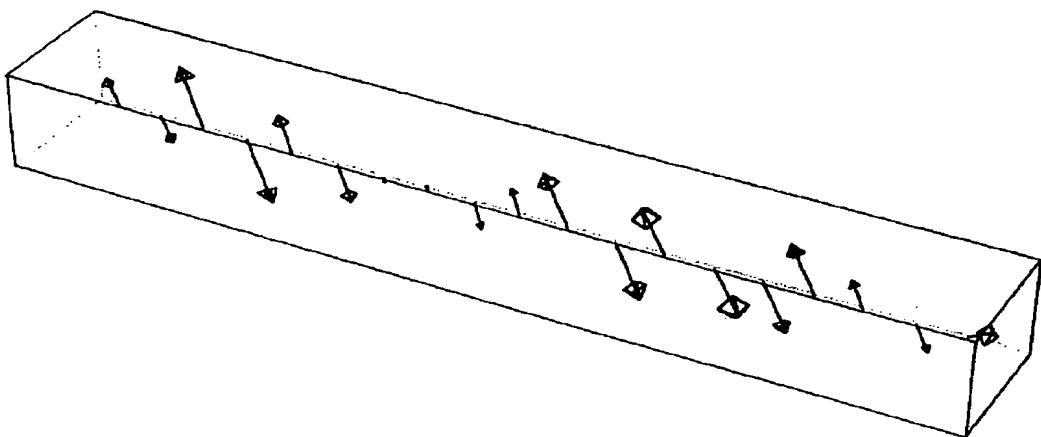
【図22】

(a)



〔図23〕

(a)



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.